Volume 193, number 1 FEBS 3172 November 1985

Hypothesis

Sequence conservation in the N-terminal domain of histone H1

L. Böhm and T.C. Mitchell

Department of Radiotherapy, University of Stellenbosch Medical School, PO Box 63, Tygerberg 7505, Republic of South Africa

Received 30 August 1985

Histone H1 *N-domain* *Conserved segment*

1. INTRODUCTION

Histone H1 is the least conserved of all histones showing a considerable degree of sequence variation between species and subfractions [1–12]. Published alignments show that the primary structure variations are strongly expressed in the N- and C-terminal part of the molecule but that sequence variation is low in a central stretch of about 80 mainly apolar residues which show a composition typical of folded globular proteins [13–16]. Spectroscopic studies on free H1 in solution and on H1 peptides have since established that all elements of secondary and tertiary structure are found in the G-domain and that the two tail domains with an abundance of basic residues lack the ability for self-organisation behaving as freely mobile extended chains [16]. From the point of sequence conservation it was expected that the conserved G-domain would perform a universal function and that species- and organ-specific functions of H1 may reside in the highly variable tail domains [2]. Addition of H1 peptides to H1-depleted chromatin and analysis of the reconstitute for protection of micrococcal nuclease sites has led to a model in which G-H1 protects 168 bp of DNA and delineates 2 full turns of superhelical DNA [19]. These experiments have furthermore shown that C-H1 is most effective in restoring salt-dependent chromatin condensation at 80 mM NaCl, a process in which N-H1 and C-H1 are much less effective [19,20].

In the search for a functional role of N-H1 in chromatin we have prepared a number of overlapping peptides of the configuration nGH1, i.e. molecules containing the complete G- and C-domain and variable parts of the N-domain [21,22]. When H1-depleted chromatin was reconstituted with these peptides it was found that addition of H1 (31–210) resulted in a distinctly better 168 bp protection than H1 (33–210) in which the number of basic residues was less [23]. When reconstitution was performed with NG-H1 and G-H1 the peptides were found to protect 168 bp nearly equally well with the NG-H1 reconstitute being more resistant to yield 145 bp length DNA [19]. The apparent requirement of the N-terminal basic cluster in the locating and anchoring of the G-domain to H1 [23] led us to suspect that operationally N- and G-domain may not be as distinct as suggested by structural considerations.

2. DISCUSSION

Close inspection of H1 primary structures shows that the N-domain comprises about 36 residues in which alanine, proline, serine and clusters of basic

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

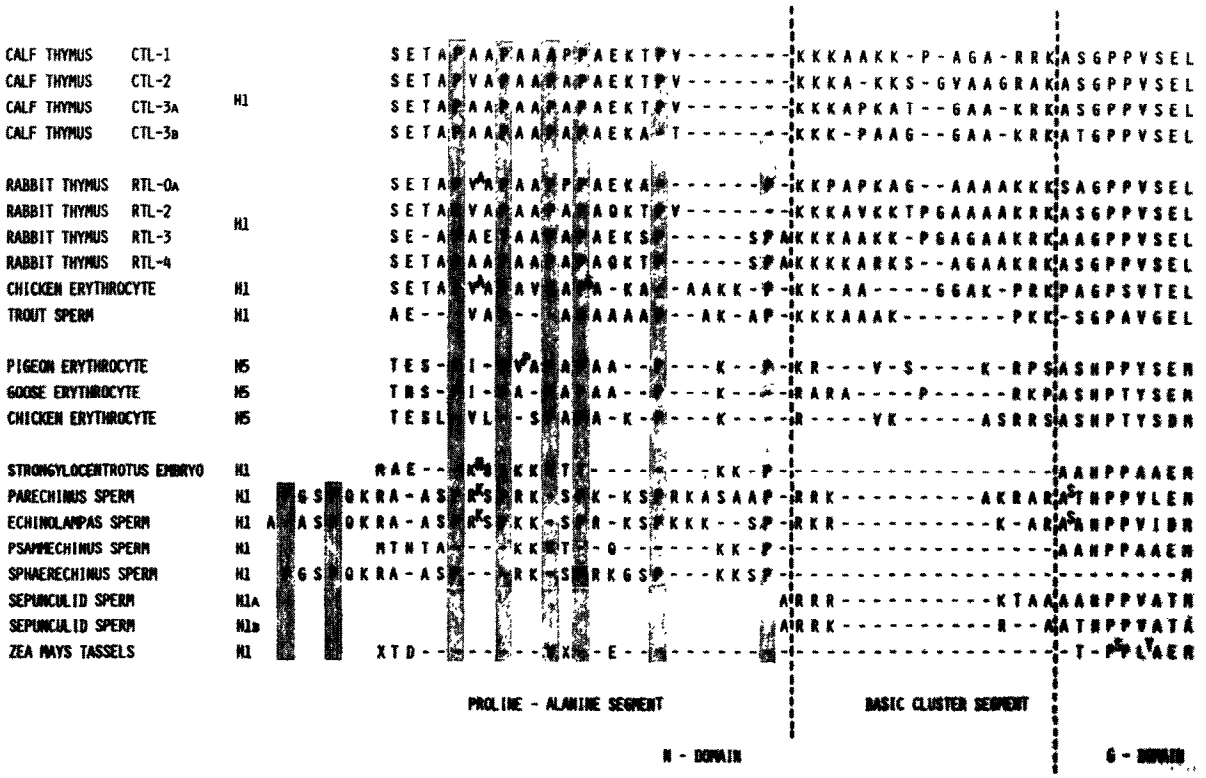

Fig.1. Alignment of the NG-region of H1 sequences using recurring prolines as points of alignment. Vertical dashed lines mark boundaries between domains or subregions within domains. (–) Deletion. Calf thymus CTL -1 [1]; calf thymus CTL-2, CTL-3a, CTL-3b (Hsiang, M. Largman, C. and Cole, R.D., personal communication); rabbit thymus RTL-2, RTL-3, RTL-4 [1]; rabbit thymus RTL-0a (Hsiang et al. personal communication); chicken erythrocyte H1 [2]; trout sperm H1 [9]; pigeon erythrocyte H5 [4]; goose erythrocyte H5 [4]; chicken erythrocyte H5 [5]; *Strongylocentrotus* embryo H1 [11]; *Parechinus sperm H1 [6,7]; Echinolampas* sperm H1 [6,7]; *Psammechinus* sperm H1 [12]; *Sphaerechinus* sperm H1 [6,7]; Sepunculid sperm H1 [24]; *Zea mays* tassels H1 [10].

residues are prominent (fig.1). The distribution of these residues is strikingly non-uniform with proline and alanine occupying the distal part of the N-domain and the clusters of basic residues occupying the part which is proximal to the G-domain. In view of the involvement of basic clusters in the anchoring of the G-domain [23] and the apolar character of the proline-alanine region it appears justifiable to consider these two segments as separate structural entities. Alignments of H1 sequences on the basis of recurring proline residues [25] show that the N-domain indeed separates into two distinct subregions, both of which are fairly well conserved (fig.1).

The hydrophobic proline-alanine segment is distinct from the highly charged proximal segment in which 90% of the basic residues are accommodated. In the calf thymus subfraction CTL-1 the basic subdomain comprises residues 20–33 showing 3 clusters of basic residues at position 20–22, 25–26 and 31–33. The second cluster at position 20–22 appears to be less stable in evolution than those at position 20–22 and 31–33 which are also present in histone H5 and in H1 species from lower organisms. A close sequence homology in the N-domain of somatic and sperm histones from echinoderms is also apparent in the alignments of De Groot et al. [28]. Histone H2B which may be considered a close relative of H1 also shows conserved prolines in the N-domain [15]. The overall impression from the alignments in fig.1 is that the extreme sequence variability of the

N-domain of histone H1 expressed in many reviews [13–16] may be an oversimplification. Alignments based on recurring prolines indeed indicate a striking degree of sequence conservation in the entire N-domain. Major primary structure differences which remain are those between vertebrates and marine intervertebrates as the latter contain large proportions of basic residues in the proline-alanine segment and seem to contain either fewer basic residues in the basic cluster segment or as evident in *Psammechinus miliaris, Strongylocentrotus, Sphaerechinus* and *Parechinus* lack this region altogether. In Sepunculid H1, on the other hand, the basic residues are located exclusively in the basic cluster segment and the proline-alanine segment is totally absent.

The alignments presented here concern the N-G region only. No attempt is made to reexamine the lack of sequence conservation in the C-domain. While the C-domain of H1 also contains proline, alanine and clusters of basic residues, it resembles the N-domain of the invertebrates in that the basic clusters are not segregated but interspersed. Attempts to rationalize alignments in the C-domain of H1 are also hampered by lack of complete sequences.

The division of the N-domain into two subregions is supported by the frequent deletion of the entire proline-alanine or the entire basic cluster segment. The distinct differences in chemistry between subdomains also support this division. In somatic histones there is good evidence that the basic cluster segment plays a role in stabilizing the chromatosome [23]. NMR studies of H1 peptides at high and low ionic strength have shown that H1 (31–210) and H1 (33–210) display 95–99% of the characteristic upfield displacement of the tyrosine resonance of intact H1 whereas H1 (42–210) only shows 20% of the displacement [26]. Elimination of the basic residues and approx. 7% of the protein mass thus resulted in a major collapse of the conformation of the whole G-domain. This suggests a close structural interrelationship between the basic cluster segment and the G-domain. The possibility that charge neutralization might induce helices in lysine- and arginine-rich regions has been considered [27] and X-ray diffraction studies of protamines indeed show a change from random coil to α-helix upon binding to tRNA for some but not all basic regions [29]. The basic cluster segment in CTL-1 under conditions of charge neutralization is predicted to be helical [30]. A close association of the basic cluster segment with the G-domain can also be inferred from a β-turn predicted at Ser 35 [26]. Zero-length crosslinkers induce N/N and N/C but no N/N dimers in chromatin [31]. This argues for a restricted rather than an extended conformation of the N-domain in chromatin.

We have presented new alignments of histone H1 sequences which strongly suggest that the N-domain of H1 consists of two distinct subregions both of which are fairly well conserved. The basic cluster segment in the N-domain may function in close association with the G-domain. The location and function of the proline-alanine segment remains unknown.

ACKNOWLEDGEMENTS

We thank Drs R.D. Cole and P. Sautiere for communicating sequence data prior to publication and Dr. C. Crane-Robinson for comments. L.B. was supported by the Medical Research Council and the Harry Crossley Fund.

REFERENCES

[1] Cole, R.D. (1977) in: The Molecular Biology of the Mammalian Genetic Apparatus (Tso, P.O.P. et al. eds) vol.1, pp.93–104, Elsevier, Amsterdam, New York.
[2] Rall, S.C. and Cole, R.D. (1971) J. Biol. Chem. 246, 7175–7190.
[3] Sugarman, B.J., Dodgson, J.B. and Engel, J.D. (1983) J. Biol. Chem. 258, 9005–9016.
[4] Yagaguchi, M., Roy, C., Dave, M. and Seligy, V. (1977) Biochem. Biophys. Res. Commun. 76, 100–106.
[5] Briand, G., Kmiecik, D., Sautiere, P., Wouters, O., Borie-Loy, O., Bizerte, G., Mazen, A. and Champagne, M. (1980) FEBS Lett. 112, 147–151.
[6] Strickland, W.N., Strickland, M. and Von Holt, C. (1982) Biochim. Biophys. Acta 700, 127–129.
[7] Strickland, W.N., Strickland, M., Von Holt, C. and Giancotti, V. (1982) Biochim. Biophys. Acta 703, 95–100.
[8] Sautiere, P. et al. (1985) Eur. J. Biochem., in press.
[9] McLeod, A.R., Wang, N.C.W. and Dixon, G.H. (1977) Eur. J. Biochem. 78, 280–291.
[10] Hurley, C.K. and Stout, J.T. (1980) Biochemistry 19, 410–416.

[11] Levy, S., Sures, J. and Kedes, L. (1982) J. Biochem. 257, 9438–9433.
[12] Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978) Cell 14, 655–671.
[13] Elgin, S.C.R. and Weintraub, H. (1975) Annu. Rev. Biochem. 44, 725–774.
[14] Cole, R.D. (1984) Anal. Biochem. 136, 24–30.
[15] Von Holt, C., Strickland, W.N., Brandt, W.F. and Strickland, M.S. (1979) FEBS Lett. 100, 201–218.
[16] Crane-Robinson, C., Staynov, D.Z. and Boltman, J.P. (1984) Comments Mol. Cell. Biophys. 2, 219–265.
[17] Hartman, P.G., Chapman, G.E., Moss, T. and Bradbury, E.M. (1977) Eur. J. Biochem. 77, 45–71.
[18] Aviles, F.J., Chapman, G.E., Kneale, G.G., Crane-Robinson, C. and Bradbury, E.M. (1978) Eur. J. Biochem. 88, 363–371.
[19] Allan, J., Hartman, P.G., Crane-Robinson, C. and Aviles, F.X. (1980) Nature 288, 675–679.
[20] Thoma, F., Losa, R. and Koller, T. (1983) J. Mol. Biol. 167, 619–640.
[21] Böhm, L., Sautiere, P., Cary, P.D. and Crane-Robinson, C. (1982) Biochem. J. 203, 577–582.
[22] Mitchell, T. and Böhm, L., unpublished.
[23] Allan, J., Mitchell, T., Harborne, N., Böhm, L. and Crane-Robinson, C. (1985) J. Mol. Biol., in press.
[24] Kmiecik, D., Sellos, D., Belaiche, D. and Sautiere, P. (1985) Eur. J. Biochem. 150, 359–370.
[25] Mitchell, T.C. (1984) MSc Thesis, University of Stellenbosch, Stellenbosch, South Africa.
[26] Mitchell, T.C., Sautiere, P., Cary, P.D., Maeder, D.L. and Böhm, L., unpublished.
[27] Maeder, D.L. (1981) The Role and Mode of Action of H1 in the H1 Mediated Condensation of Nucleosomal Chromatin, PhD Thesis, University Cape Town.
[28] De Groot, P., Strickland, W.N., Brandt, W.F. and Von Holt, C. (1983) Biochim. Biophys. Acta 747, 276–283.
[29] Wade Warrant, R. and Kim, S.H. (1978) Nature 271, 130–135.
[30] Lewis, P.N. and Bradbury, E.M. (1974) Biochim. Biophys. Acta 336, 153–164.
[31] Ring, D. and Cole, R.D. (1983) J. Biol. Chem. 258, 15361–15364.